



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
[www.uspto.gov](http://www.uspto.gov)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/703,174	10/31/2000	Charu C. Aggarwal	Y0R920000430US1	7445
7590	05/02/2011			EXAMINER
William E Lewis				HILLERY, NATHAN
Ryan Mason & Lewis LLP				
90 Forest Avenue			ART UNIT	PAPER NUMBER
Locust Valley, NY 11560			2177	
			MAIL DATE	DELIVERY MODE
			05/02/2011	PAPER

**Please find below and/or attached an Office communication concerning this application or proceeding.**

The time period for reply, if any, is set in the attached communication.

UNITED STATES PATENT AND TRADEMARK OFFICE

---

BEFORE THE BOARD OF PATENT APPEALS  
AND INTERFERENCES

---

*Ex parte* CHARU C. AGGARWAL and PHILIP SHI-LUNG YU

---

Appeal 2009-004329  
Application 09/703,174  
Technology Center 2100

---

*Before* JOSEPH L. DIXON, LANCE LEONARD BARRY, and  
CAROLYN D. THOMAS, *Administrative Patent Judges*.

THOMAS, *Administrative Patent Judge*.

DECISION ON APPEAL

### STATEMENT OF THE CASE

Appellants appeal under 35 U.S.C. § 134(a) from a final rejection of claims 1-27. We have jurisdiction under 35 U.S.C. § 6(b).

We REVERSE.

According to Appellants, the invention relates “to methods and apparatus for performing large scale collection of web pages from the world wide web [WWW]” (Spec. 1: 5-6). More particularly, the invention involves intelligent crawling techniques which “provide a crawler mechanism which is capable of learning as it crawls in order to focus the search for documents on the information network being explored, e.g., [the] world wide web” (*see* Abstract).

Claim 1 is illustrative:

1. A computer-based method of performing document retrieval in accordance with an information network, the method comprising the steps of:

initially retrieving one or more documents from the information network that satisfy a user-defined predicate, wherein the initial document retrieval operation is performed without assuming a specific model of a linkage structure such that the initial document retrieval operation retrieves the one or more documents without assuming that a relationship exists between a feature of a first one of the one or more documents and a feature of at least another one of the one or more documents that links to the first one;

collecting at least a set of aggregate statistical information and a set of predicate-specific statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed; and

using the collected statistical information to automatically determine further document retrieval operations to be performed in accordance with the information network, wherein the statistical information using step further comprises learning a linkage structure from at least a portion of the collected statistical information with each successive document retrieval operation such that the learned linkage structure is available for use in performing subsequent document retrieval operations requested by a user.

*Rejections*

R1: Claims 1-8, 10-17 and 19-26 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over SOUMEN CHAKRABARTI, FOCUSED CRAWLING: A NEW APPROACH TO TOPIC-SPECIFIC WEB RESOURCE DISCOVERY (hereafter “Chakrabarti 1”) and Chaudhuri (US 6,529,901 B1, Mar. 4, 2003).

R2: Claims 9, 18 and 27 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Chakrabarti 1, Chaudhuri, and SOUMEN CHAKRABARTI, DISTRIBUTED HYPertext RESOURCE DISCOVERY THROUGH EXAMPLES 375-386 (Proceedings of the 25th VLDB Conference 1999) (hereafter “Chakrabarti 2”).

**FINDINGS OF FACT (FF)**

*Appellants' Specification*

1. Appellants' Specification discloses:

The aggregate statistical information contains two kinds of information:

- (1) The number of times each word has occurred during the entire process of crawling.

(2) The number of times that each token in any URL has occurred during the entire process of crawling. (Spec. 9:24 - 10:2).

*Chaudhuri Reference*

2. Chaudhuri discloses:

The MNSA [Magic Number Sensitivity Analysis] technique for determining if the existing set of statistics contains an essential set of statistics should be qualified as follows .... Third, although for SPJ [Select-Project-Join] queries MNSA ensures that an essential set is included among the statistics, it is necessary to extend the method beyond simple queries. Aggregation (GROUP BY or SELECT DISTINCT) clauses can be handled by associating a selectivity variable that indicates the fraction of rows in the table with distinct values of the column(s) in the clause. For example, a value of 0.01 for such a selectivity variable for the clause GROUP BY ProductName implies that the number of distinct values of ProductName is 1% of the number of the rows in the table (see col. 19, ll. 35-55).

ANALYSIS

*Claims 1-27*

Our representative claim, claim 1, recites, *inter alia*, “*collecting at least a set of aggregate statistical information.*” Independent claims 10 and 19 recite similar limitations. Thus, the scope of each of the independent claims includes collecting aggregate statistical information.

Issue: Did the Examiner err in finding that the combination of Chakrabarti 1 and Chaudhuri teaches or suggests “collecting at least a set of aggregate statistical information,” as recited in representative claim 1?

The Examiner found that “Chaudhuri et al. teach gathering statistics by handling aggregation clauses which is equivalent to the claimed set of aggregate statistical information.” (Ans. 11.)

Appellants contend that Chaudhuri’s “disclosure of a manner in which GROUP BY or SELECT DISTINCT clauses may be handled fails to teach or suggest a limitation directed toward collecting a set of information maintained for all retrieved documents.” (App. Br. 7.)

Appellants rely upon the fact that their Specification discloses that “[t]he aggregate statistical information is maintained on all the retrieved web pages” (FF 1a). We find that this is merely a “requirement” statement for *maintaining* information rather than a “definition” for “aggregate statistical information.” During examination, claims are to be given their broadest reasonable interpretation consistent with the specification, and the language should be read in light of the specification as it would be interpreted by one of ordinary skill in the art. *In re Amer. Acad. of Sci. Tech Ctr.*, 367 F.3d 1359, 1364 (Fed. Cir. 2004) (citations omitted). The Office must apply the broadest reasonable meaning to the claim language, taking into account any definitions presented in the specification. *Id.* (citations omitted). Here, Appellants’ Specification defines *aggregate statistical information* as information that *contains two kinds of information*, i.e., (1) the number of times each word has occurred during the entire process of crawling and (2) the number of times that each token in any URL has occurred during the entire process of crawling (see FF 1). In other words, the claimed “aggregate statistical information” includes the number of times each word

has occurred and the number of times that each token in any URL has occurred, during the entire crawling process. The Examiner found that Chaudhuri's "aggregation clauses" meet the limitations set forth above (Ans. 10.). We disagree.

While Chaudhuri certainly suggests aggregating information using the disclosed aggregation clauses (*see* FF 2), the Examiner has not shown, and we do not readily find, that Chaudhuri discloses that the statistics gathered by the GROUP and SELECT DISTINCT clauses includes an accounting of (1) the number of times a word occurs and (2) the number of times a token in any URL occurs during the MNSA technique, i.e., crawling procedure. Instead, the columns cited by the Examiner merely shows that Chaudhuri associates a selectivity variable with distinct values of the columns (FF 2). However, the Examiner has not shown how such an association is equivalent to an accounting of the number of times a word occurs and a token occurs, which is a requirement of collecting aggregate statistical information as set forth in claim 1.

Since we agree with at least one of the arguments advanced by Appellants, we need not reach the merits of Appellants' other arguments. It follows that Appellants have shown that the Examiner erred in finding that the combination of Chakrabarti 1 and Chaudhuri renders claims 1-27 unpatentable.

Thus, we find that the Examiner has erred in finding that Chaudhuri 1 teaches or suggests "aggregate statistical information," as recited in representative claim 1. Independent claims 10 and 19 are commensurate in scope with the argued limitation. Accordingly, we reverse the Examiner's

Appeal 2009-004329  
Application 09/703,174

§103 rejection of representative claims 1 and of claims 2-27, which contain the same deficiency and stand therewith.

#### DECISION

The Examiner's rejection of claims 1-27 under 35 U.S.C. § 103(a) is reversed.

**REVERSED**

ke